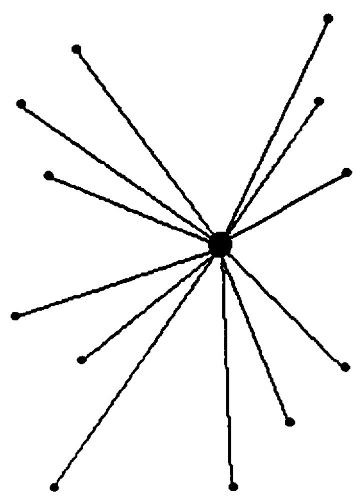
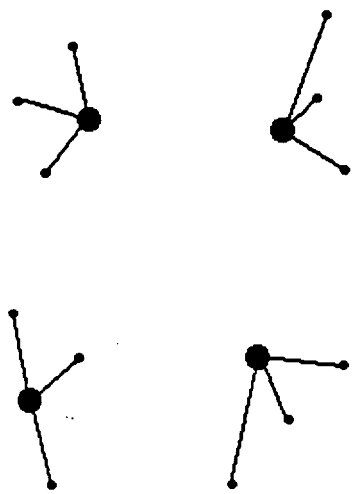
2.2.2 Mirroring Server Sites

Mirroring is a technique used by many web site providers to handle large loads (hits). In mirroring, an organization's web site is duplicated on a number of independent servers, called mirror sites. Routing and other network mechanisms are used to balance the traffic load across these servers, usually sending traffic to the nearest server. Thus, a number of relatively small and inexpensive servers operating independently can service the needs of many web site users. At certain intervals, the data in all of web sites are updated so that each web site is identical to its siblings.



Single Server

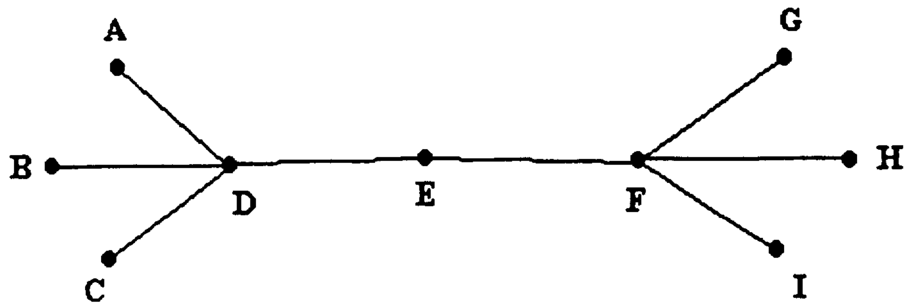


Mirrored Servers

Usually, the mirror sites are located at geographically diverse locations. This provides increased reliability across the network of servers. If one server fails, the web site is not out of business - traffic can simply be routed to the nearest (or least loaded) neighbor server. Additionally, the geographic dispersion of web servers naturally balances the load among the servers.

From the network side, total network traffic is reduced with a mirroring system. Instead of traffic having to traverse the network between the single web site and the user, the user's requests are handled by a closer web server. Traffic is regional rather than network wide. Or, to put it another way, everyone is closer to his or her server.

Consider the example in the simple nine-node network below. For simplicity, assume that the same number of users generating the same amount of web traffic (T) exist at each node of the network resulting in a total of 9T traffic. If a single web server is located at node E, then each user's traffic must be exchanged between the server site and the user. Some of the traffic must travel over more than one (two in this case) links to get to the server. Traffic originating at E does not have to travel over any links. The total link traffic (i.e. the sum of the traffic carried by all links) is 14T.



If two mirror servers are located at nodes D and F, the resulting network link traffic is only 7T.

The example above assumed equal users and traffic. Practically, the benefits of mirroring are even more dramatic when these traffic and users are concentrated. The table below lists the population of the states in terms of numbers and percentages in order of decreasing population. By cleverly selecting the locations of mirror servers, a small number of servers can service locally (i.e. intra-state) a large proportion of the total traffic. A single web site located in the largest state (California) has an intra/interstate ratio of 12%. Only nine servers appropriately located will service over half of the population as local traffic.

State	Population ('000s)	%	Cumulative %
California	32,268	12.06%	12.06%
Texas	19,439	7.26%	19.32%
New York	18,137	6.78%	26.10%
Florida	14,654	5.48%	31.57%
Pennsylvania	12,020	4.49%	36.06%
Illinois	11,896	4.44%	40.51%
Ohio	11,186	4.18%	44.69%
Michigan	9,774	3.65%	48.34%
New Jersey	8,053	3.01%	51.35%
Georgia	7,486	2.80%	54.15%
North Carolina	7,425	2.77%	56.92%
Virginia	6,734	2.52%	59.44%
Massachusetts	6,118	2.29%	61.72%
Indiana	5,864	2.19%	63.91%

Washington	5,610	2.10%	66.01%
Missouri	5,402	2.02%	68.03%
Tennessee	5,368	2.01%	70.03%
Wisconsin	5,170	1.93%	71.97%
Maryland	5,094	1.90%	73.87%
Minnesota	4,686	1.75%	75.62%
Arizona	4,555	1.70%	77.32%
Louisiana	4,352	1.63%	78.95%
Alabama	4,319	1.61%	80.56%
Kentucky	3,908	1.46%	82.02%
Colorado	3,892	1.45%	83.48%
South Carolina	3,760	1.40%	84.88%
Oklahoma	3,317	1.24%	86.12%
Connecticut	3,270	1.22%	87.34%
Oregon	3,243	1.21%	88.55%
Iowa	2,852	1.07%	89.62%
Mississippi	2,730	1.02%	90.64%
Kansas	2,595	0.97%	91.61%
Arkansas	2,522	0.94%	92.55%
Utah	2,060	0.77%	93.32%
West Virginia	1,816	0.68%	94.00%
New Mexico	1,730	0.65%	94.65%
Nevada	1,677	0.63%	95.27%
Nebraska	1,657	0.62%	95.89%
Maine	1,242	0.46%	96.36%
Idaho	1,210	0.45%	96.81%
Hawaii	1,187	0.44%	97.25%
New Hampshire	1,173	0.44%	97.69%
Rhode Island	987	0.37%	98.06%
Montana	879	0.33%	98.39%
South Dakota	738	0.28%	98.66%
Delaware	731	0.27%	98.94%
North Dakota	641	0.24%	99.18%
Alaska	609	0.23%	99.40%
Vermont	589	0.22%	99.62%
District of Columbia	529	0.20%	99.82%
Wyoming	480	0.18%	100.00%
Totals	267,634	100.00%	

For the reasons discussed above, mirroring is rapidly becoming a popular practice by web site providers. Many large web providers (e.g. Microsoft, Canon, etc.) have several mirror sites. Mirroring's benefits are not limited to web service. Mirroring will become an important architecture for other large Internet based applications such as electronic commerce.

Music Ltd. lists mirror servers in 35 states in its web site. Netscape Communications currently lists mirror servers in nine states. In addition, Netscape has a formal program for establishing new mirror sites. Details of this program can be found at their web site at the URL: http://home.de.netscape.com/comprod/mirror/mirror.html.

2.2.3 Internet Web Caching

Internet caching involves storing copies of Web pages or other content (databases) on servers that are geographically closer to end-users. Although architectures vary, the concept is to have a cache server act as a middleman between the end user and the home servers of web sites. Located physically closer to the user, the cache server provides a copy of the requested site's pages, thereby relieving the home site from having to process the request, and more importantly, eliminating costly and time consuming Internet back-bone transmission. Studies and surveys of user Web searching behavior consistently show that users tend to visit the same sites repeatedly across Web sessions, and that certain sites are popular across multiple users, thereby making caching an efficient technique. In fact, Zipf's Law has been applied to Web sites, which shows that request frequency is inversely proportional to a site's popularity ranking. From a user's perspective, cached content loads faster but could be outdated. From the network provider's point of view, caching decreases the amount of duplicate traffic on their networks. From the content provider's point of view, it decreases the load on their web servers but at the same time eliminates accurate "hit" statistics and increases worry about outdated information. From a user's perspective, one metric for evaluating an ISP performance is how quickly web pages can be loaded. Adding additional bandwidth only helps marginally because the main factors are 1) Distance between users and servers, and 2) Congestion on heavily networks and popular server sites. Mirroring discussed above is one solution but caching web pages is an addition help that also brings data closer to users. Caches will reside throughout networks, concealing the Web's slowness from clusters of users. ISPs and backbone providers will provide caches at every point of presence and enterprises such as campus networks and company branch offices. Forrester Research estimates that network caches will become as ubiquitous as IP routers: "By 1999, all Fortune 1000 companies, will depend on hardened Web caches to keep their intranets running smoothly." ("Why Caching Matters"). America Online (AOL), Prodigy, MSN, and Earthlink are all currently deploying national networks of cache servers. Both Netscape and Internet Explorer are capable of using network caches. Both are capable of channeling content to their local machine's cache for local processing (after checking for any changes in the content). The statistics on ISP caching shows a minimum of a 30% to 40% hit rate with several showing a 50% to 60% hit rate. (http://ircache.nlanr.net/cache/FAQ/ircache-faq-2.html). Hit rate represents the percentage of times a Web page is found in the ISP's web cache and does not require using the Internet backbone to retrieve the requested page. The Web site at http://ircache.nlanr.net/Cache/FAQ/ircache-faq-2.html in section 2.9 "How Effective is Web Caching" provided the above statistics. In addition, actual hit ratio statistics from NLANR Web caching can be found at ircache.nlanr.net/Cache/Statistics/Reports. The hit ratio is a function of the number of users, cache size, and the refresh parameters. In any case, as Internet Web caching increases, the need to utilize interstate resources across the

Internet backbone will continue to decrease significantly. For a list of Web caching products and services see the URL: http://ircache.nlanr.net/cache/FAQ/ircache-faq-9.html

2.2.4 Dial Access Speeds Vs ISP Internet Access Speeds

ISPs provide Internet access to their customers by providing connectivity to the Internet and protocol processing. The ISP has large numbers of ports accessible by customers through dial-up or leased lines. These access ports are usually attached to modem banks that operate at speeds typical of the modems provided by personal computer hardware vendors, generally 14.4Kbps to 56Kbps. Dial-up access is accomplished through the public switched telephone network. The ISPs attach to the Internet at Internet POPs through high speed leased lines generally PRI or higher. Thus, the nature of the access is circuit or connection oriented between the customer and the ISP, and connectionless between the ISP and Internet.

When holding time is used as a basis to determine the ratio of local to Internet bound traffic, a natural compression occurs. The customer is transferring data to the ISP at low speeds and the ISP is placing that traffic on the Internet at high speeds. For example, if a customer transfers one second's worth of traffic to the ISP at 14.4Kbps, that traffic will occupy less than one one-hundredth of a second on a PRI access line. The table below gives the relative time compressions for the typical speeds at which dial-up customers access ISPs versus ISP/Internet access lines.

Customer Access Modem Speed Kbps	Internet Access Speed (Mbps) 1.5	45
9.6	0.0064	0.000213
14.4	0.0096	0.00032
28.8	0.0192	0.00064
56.0	0.037333	0.001244

The table illustrates that, because of time compression, a dial-up customer uses a significantly smaller proportion of the ISPs Internet access line if holding time is used as the basis of measurement. Further, speeds of existing dial-up local access lines have almost reached their theoretical capacity limit. The speeds of the class of access lines that the ISPs use to access the Internet are virtually unbounded. This suggests that the ratio of local to total holding time will become even larger in the future.

Portal Effects

Large ISPs, in addition to providing Internet access, host web sites, create content and engage in other related activities. Their position as gatekeeper or "portal" as it is referred gives them influence over the default or initial and, in some cases, even ultimate content to which a customer is directed. In other words, it behooves an ISP to direct its customers to the sites of its other customers. This serves to keep as much of the traffic as possible within the ISP's business. Indeed, this power to influence is one of the rationales for the Justice Department's bringing an antitrust action against Microsoft.

2.3 Internet Usage Characteristics

There have been numerous studies performed on Internet usage and they have general agreement in most areas. The Pew Research Center surveyed 1,003 on-line Internet users in October of 1996 and found that 77% of the users receive or send e-mail at least once every few weeks and 26% use e-mail every day, while 19% use it 3 to 5 times a week. On average, e-mail users send 6 messages and receive 12 messages at each Internet session, but a Pitney Bowes study found that office workers send and receive 30 e-mails per day. AOL handles 34 million messages per day for 13 million subscribers, and Forrester Research estimates that about 500 million e-mails are delivered each day. 68% of Internet access is done from the home; 47% go on-line from work; and 24% go on-line from both home and work. The average speed with which users access the Internet is at 28,800 baud, with the next most popular speed being 14,400 baud. Only 7% use a modem with a speed of 9,600 baud or less. The percentage of users using the Internet who come from the suburbs or small towns is approximately equal at 31% and 32%, respectively, while large cities accounted for 22% of use, and rural areas accounted for 14%. The Third MIDS Internet Survey shows that in over 1,000 survey responses, students account for the majority (59.7%) of Internet access, followed by network users of the Internet such as staff in an organization (34.2%) and faculty in a university (3.5%). Approximately 14% of organizations' access to their ISPs was via a dialup mode. This study also showed 49.7% of the respondents' computers mailed out from their workstations, with the most frequently used speed of 64Kbps (22%) followed by 28.8Kbps (17.1%) and 14.4Kbps (12.5%). 1.544Mbps was used by 10.2% of the organizations responding. E-mail was by far the most heavily used application. Web server statistics by several surveys are in close agreement that the .edu domain accounts for the majority of accesses to servers, with the .com domain gaining rapidly. This correlates well with the fact that students and faculty have historically accounted for the majority of use. A survey at www.erg.cuhk.edu.hk shows that .com domains account for over 3.9 million hosts followed by the .edu domain with over 2.65 million hosts. Third in their list of hosts was the .net domain with over 1.5 million. The most frequently used host name was WWW followed by HOST, MAIL, and FTP. This provides an indication of the expected services to be offered by these hosts.

Login and usage statistics from a University of Wisconsin study of 58 selected users between 1/22/1996 to 5/17/1997 (4 months) showed the average number of logins was 17.47 and sessions averaged 22.6 minutes with 51% being between 1 and 10 minutes. (see www.wisc.edu/learntech/firstclass/logstats).

A survey (89 responses) done by the Oil & Gas Industry (http://209.49.75.165/internetresults.htm) showed their members used the Internet in the following ways:

Main Use	Percent of Use
Collaboration	16.9
Competitive Intelligence	23.6
E-mail	47.2
Employment	22.5
File Transfer	25.8
General Info	49.4
Industry Data	31.5
Industry Discussion	8.9
Industry News	32.6
Market Research	11.2
Product Info	29.2
Product Purchase	12.4
recreation/Leisure	22.5
Research	29.2
Software Download	30.3
Travel Reservations	13.5

Oil & Gas Industry Survey – Internet Usage

The above chart shows that web browsing and e-mail are the two major uses of the Internet. The question asked was "What do you mainly use the Internet for?" and people could respond with multiple uses. 47.2% of the respondents indicated that they used it for e-mail while the rest of the responses indicate that it was used for browsing the Web for information, file transfer, and software downloads. This corresponds with surveys done for educational use and home use of the Internet. For instance, Morgan Stanley estimates that the number of e-mail users on a daily basis was 35 million in 1995 and will be 200 million in 2000. They also estimate that the number of web users was 9 million in 1995 and will be 152 million in 2000 (See www.cyberatlas.com/demographics.html). The most intensive use of communication bandwidth resources in a short time period is file transfer and software downloads, but they occur much less frequently than e-mail and web browsing.

Results of other Internet usage surveys are available upon request.

2.4 Methods Used

To provide evidence related to the problem required obtaining data about how much time was spent using interstate telecommunication resources versus using local telecommunication resources for various types of Internet uses such as e-mail and web searching. Some general data was derived from statistics made available on the Internet by ISPs and various web sites, while other data was collected by having users keep a log of Internet sessions and perform transmission timing experiments based on the Internet session. User data was logged on a paper or spreadsheet using the survey forms shown in Appendix A. The timing experiments were used to determine average transmission time from the user site to an Interstate destination or origination Internet site. Some traffic data was also collected from ISPs to compare session time and utilization of local and long distance communication resources on the Internet derived from the user-based experiments. The methodology builds a worse case scenario to provide evidence related to the problem. This was done to compensate for the fact that user logged data are approximations.

The methodology is based on a simplified model of how Internet client/server applications function relative to the 10% mixed traffic rule. The critical variables in the federal regulation being tested are **total holding time, long distance holding time and their ratio. Total holding time in this study is equivalent to session time. Session time is defined as the time that elapses between the start of an e-mail or web session and its termination by the user. An e-mail or web session is assumed to start when the e-mail program for web browser is requested to execute by the user. Likewise, an e-mail or Web Browser session is terminated when the user clicks the exit option. Subjects were asked to explicitly initiate and terminate each session in this manner.** These times were recorded by users to the nearest minute and converted to milliseconds. We excluded individuals that keep an e-mail session open all day. This practice is more common than might be expected, but would bias the study in favor of less than 10 percent. **Long distance interstate holding time is defined as the amount of time that a connection is held between the user's client process and an out-of-state site (server process) on the Internet for completing a transaction. A transaction includes a request from a browser or e-mail program and the response from a server. Interstate long distance holding time is calculated as the product of the average number of bytes in an Interstate long distance message or web page and the average time in milliseconds to transmit a byte from the long distance interstate sites as determined by a "PING" experiment as explained below. A Web page is defined as all the data delivered to a web browser in response to a single request. This data was measured in bytes and was averaged across all Web pages retrieved. This data is available in a browser's web page cache file and can be examined by the user.** The bytes (length) in an e-mail message were simply counted manually or the message was saved on disk where the length could be determined automatically.

Data for the study was self-reported by a group of 114 actual Internet users who each logged 1 to 3 e-mail and web browsing sessions selected over a period of several weeks. The type of users ranged from full time graduate students and professional workers to homemakers. The initial set of users was 86 students in a graduate class in Information Science at the University of Pittsburgh who also recruited 28 other participants. The professionals in the study hold a diverse set of positions in a variety of institutional settings. The majority of the users could be considered frequent users as well as knowledgeable about the Internet since the majority were graduate students in an Information Technology intensive program and most worked in technology intensive jobs. These users represent a worst case scenario since they make extensive use of e-mail attachments, visit foreign sites, and utilize graphic images extensively. Additionally, many students are from foreign countries that make extensive use of e-mail to communicate with their family and friends abroad. The student users were asked to recruit non-student participants from friends, colleagues, and family. Internet accesses included accesses from an organizational setting with dialup and direct connections and dial-up from home. The participants were asked to maintain an accurate log of their Internet e-mail and web-browsing activities from which traffic load (bytes of data sent and received) and timing (transmission and session duration) data was collected and computed for the equivalent of what would be a long distance (interstate) or local telephone call. The subjects were trained on how to gather and record the data during a 30-minute demonstration. The participants used a paper form to log each of their Internet sessions but were permitted to e-mail it as a text document or spreadsheet to the investigators.

Participants were asked to be as accurate as possible in recording the start and end times for an e-mail or web session. If they forgot to record either time, they were asked to simply discard the session and record a new one. If the data was received without such times it was discarded. The most difficult part of the session logging was identifying local sites from interstate sites. An interstate site was considered any site outside the state of Pennsylvania. For e-mail, this was not very difficult since most e-mail is from known individuals with known locations, but for web searching and surfing, the problem was much more difficult. Participants were required to log the URL of each page they accessed. These are actually available in the browser cache and can be examined after a session is competed. The participants were then required to determine whether the location of the server for the URL was instate or out-of-state. For international sites this was simple because the URL indicates the country in the last part of the URL. For many sites in the USA, the URL is also self-evident, such as www.pitt.edu at IP address 136.142.140.120 at the University of Pittsburgh in Pittsburgh, Pa. For sites that were not self-evident, the physical address of the organization owning the server was used. Physical addresses were determined by first looking at the home page to see if an address was provided. Then a search was made on the web to see if a physical address could be found. The use of sites such as www.stocksite.com for commercial sites to get an address was very successful. When two addresses were available (1 in-state, 1 out-out-state), the out-of-state address was used. When no physical address could be determined, the site was discarded. If an interstate site was not PINGable, it was discarded from consideration for timing purposes.

Participants were required to log the number of Web pages they accessed per site and compute the average number of bytes per page by using the browser's cache information. Both of these are important in computing interstate holding time. The average bytes per page in the study is over 20,000 for interstate sites and 22,000 for intra-state sites. These differed slightly for dial access and direct access users. A Web page typically has a combination of text and one or more images but other Web applications may download data files, executables, movies, audio clips, etc.

The web site at ircache.nlanr.net/Cache/Learn/learn-1.html provides average document sizes for different document types as collected during 1 week in April 1997 and is shown below:

Document Types	Average (mean) Size in Bytes
Image	7931
HTML	6419
Directory	6659
Lookup	2420
Query	5308
Applet	4201
Text	26954
Audio	105648
Executable	1206714
Movie	864532
Software	4450
PDF	287994
PostScript	326692
VRML	24371
ISMAP	1250

Over 50% of the total number of bytes transmitted were HTML and images, while the percent of HTML and Image documents transmitted was 84%. Thus, the survey average of 20 to 22 thousand bytes per page is reasonable.

They were also asked to estimate how much time that they spent reading the data they received on a web page.

The TCP/IP PING utility was used to get data on the transmission time to and from the Interstate e-mail and web sites involved. These Internet activities were timed in terms of local resource holding time (session time) versus Interstate long distance holding time for the transport of e-mail and web page data (including text, image, video, audio data, etc.) based on the client/server model. The time for transmitting bytes to/from Interstate sites was longer on average than for Intra-state sites as might be expected. The PING time includes network virtual circuit setup and termination time as well as transport time for TCP and transport time for UDP. **This PING time does not include processing time at a server site for retrieving and processing data. On the other hand, it does include transport time between the end-user and the ISP at both the server and client sites which constitute local connect time and is the slowest part of the transmission pipeline since dial-up users are typically transmitting at 28.8 Kbps while the Internet backbone is transmitting at 45 Mbps and higher. It was not feasible in this study to isolate the processing time at a server or the local transmission portion of the transport. It is suspected that these times may on average offset each other.** The browser cache information was used to determine the URLs of sites visited for PING timing experiments. The percent of cache hits (31.25%) was determined by averaging cache hit-rate data from Web sites such as ircache.nlanr.net/Cache/Statistics that report such data. This was done by randomly selecting 30 days to use to get sample points. No consideration was given to the rapidly developing mirroring of sites that is being implemented by many of the most popular web sites due to a lack of empirical data. The data was analyzed using descriptive statistics such as the arithmetic mean, but more importantly, the ratio of local holding time (session time) to interstate resource holding time. The Interstate long distance holding time for connection and transport includes the transport from the user workstation to the ISP, as well as from the ISP to the Internet POP. Since these transport facilities are typically much slower than the Internet backbone, there is a bias toward longer Interstate long distance holding time than is actually true. There is no simple way to subtract these times from the total transmission time in this experiment. Based on this data and its analysis, evidence is presented relative to the mixed traffic 10% rule for these applications. The log sheet used by the subjects in this study is contained in Appendix A.

Although empirical data for other types of Internet based applications is not available from this study, we can use the results of the e-mail and Web browsing survey to make some inferences related to applications such as electronic commerce since they have similar characteristics to e-mail and web browsing.

3.0 Results of the Study

3.1 E-mail

E-mail is the most highly used application on the Internet. An E-mail transaction has several phases as outlined below:

- Connect to a local ISP via dialup or direct access
- Download messages to the local desktop machine from the ISP mail server
- Get a list of waiting messages
- Select a waiting message to read
- Read messages and viewing or saving attachments
- Reply to messages and composing new messages
  - Addressing the message (local or long distance)
  - Composing the message text
  - Appending attachments
  - Transmission of messages to the local ISP from the user's local machine
  - Queuing of the message at the ISP mail server
  - Placing the message in a local mailbox if destination is same ISP
  - Performing a URL IP address resolution
  - Transmitting the message as packets over the Internet towards its destination (Same state or Interstate)
  - Intermediate processing at routers and the destination host mail sever

- Saving, replying (sending) or deleting messages at the local machine
- Disconnect from the local ISP

Since another Internet user sends the e-mail received, either receiving time or sending time as calculated from the number of Interstate Long Distance (LD) bytes received or sent in messages should be used as holding time but not both. Otherwise, an e-mail message would be counted twice in the calculation (at both the sender and receiver ends). We have opted to use the Interstate LD computed receive time since the Interstate LD holding time is the larger of the two times in our sample. The Interstate (LD) holding time is calculated as the product of the average number of Interstate LD bytes received and the average time to transmit a byte to the PINGable sites. The total Interstate LD bytes received is the product of the average number of Interstate LD messages received per e-mail session and the average message length in bytes of a received e-mail message ,plus the product of the average number of attachments per session and the average size of an attachment. The transmission time is the average of the PING time to all Interstate LD sites from which e-mail was received.

Two hundred thirty one (231) e-mail sessions with 1935 e-mail messages received and 260 Interstate PINGable sites were analyzed with the following results:

Variable	**Average Value (Arithmetic Mean)**
Number Local Messages	4.80
Number Long Distance (LD) Messages	4.26
Number Interstate LD Messages	3.47
Bytes per Message	907.73
Number Attachments per Session	.995
Number Attachments per Message	.110
Bytes per Attachment	27,968.26
Number Responses to Messages per Session	2.63
Number Bytes per Response	578.96
Number Responses to Local Messages	1.59
Number Responses to LD Messages	1.20
Time to Compose a Response	4.33 Min
Number New Messages sent/session	1.29
Bytes per New Message sent	842.37
Time to Compose a New Message	6.12 Min
Session Length – Total Holding time	31.74 Min (1,904,299 ms)
E-mail Interstate Site PING Time (Round Trip)	274.49 ms
Bytes per PING	49.6
Byte Transmit Time	2.767 ms
Total Bytes Received per e-mail Session	36,063.40
Total LD Interstate Bytes Received per Session	30,988.94
Total LD Interstate Holding Time	85,746.40 ms
Total Holding (Session) Time	1,904,299 ms

The ratio of Total Holding time (session time) to Long Distance Holding Time is 85,746.40 ms / 1,904,299 ms = .045 or 4.5 percent. This would indicate that the holding time for long distance resources is less than 10 percent of the total holding time (session time). The PING time used in computing transmission time includes network overhead. Network overhead includes the hand-shaking protocol and network control messages between the client and server to establish a session and any retransmissions caused by errors of any type.

3.2 World Wide Web (WWW) Searching and other Web-based Applications

Searching the World Wide Web (WWW) is composed of the following phases using a client/server mode:

- A connection is made to an ISP
- A URL is specified to the client browser (this may be a default URL at startup)
- The Client browser looks in its local cache to see if the requested page is stored and if so retrieves and displays it without any need to use the network facilities.
- If the page is not cached, then the client browser performs a Domain Name Resolution Request to get an IP address for the URL
- The individual end user's computer or local network, or an ISP using a caching web server, will attempt to find the requested page, and if found send it to the browser without using the Internet.
- If there is no caching by the local network or an ISP, then the client browser establishes a connection over the Internet with the IP addressed destination server site
- The client browser transmits its request to the destination server site
- The WWW server at the destination site processes the client browser request
- An HTML page with possible links to text, image, audio, and video are transmitted as series of packets to the client requestor site
- The server terminates the connection with the client requestor
- The ISP or organization server may cache the pages locally
- The client browser processes the received HTML page and associated data for display and/or sound
- The client caches the pages locally for speeding up another reference to the page
- The client browser uses the preceding steps to follow links (HREF) to retrieve referenced pages
- The user views and processes the received data
- The client browser waits for the user to initiate the next request and the above steps are performed as applicable

This process is different from e-mail since the user is typically performing a search task to find information as opposed to retrieving mail locally. It is possible that a user's mail server could be located at an interstate location relative to the user's location, but this would be an exception and not the rule. It would most likely occur via home use since organizations tend to have local mail server resources. Even if the main server were at an interstate site, the transfer time for mail to the local machine occurs at file transfer speeds since all messages are downloaded to the desktop mail client and responding using mostly local resources. Although the browser does get some data from the local client to transmit to a remote server, this data is minimal in nearly every case comprising a few words for search arguments. The transmission of data from the browser to a server is NOT included in the calculations below and is assumed to have very little impact in any case. One complicating factor for computing data transmitted from remote sites to a client (browser) is that browsers and servers cache the Web page data most recently accessed. This decreases the need to revisit sites for which the Web page data is in the local cache. Web cache statistics were gathered from several web sites such as the ones found at ircache.nlanr.net/Cache/Statistics and netstat.uu.se/Stat/Squid. The question is how much time is spent using Internet interstate facilities to retrieve updated or new Web pages in comparison to the total time spent connected to an ISP.

One hundred seventy two (172) web-browsing sessions with 356 PINGable sites were analyzed with the following results:

Variable	Average (Arithmetic Mean)
Number of Local Sites Visited	2.63
Number Long Distant (LD) Sites Visited	4.97
Number LD Intra-State Sites Visited	1.01
Number LD Inter-State Sites Visited	4.01
Number Pages Retrieved per Site	3.14
Number Bytes per Page from LD Sites	20,930.42
Number Bytes per Page from Local Sites	22,678.99
Time to View/Read All Pages	24.89 Min
Session Length (Total Holding Time)	66.58 Min (3,994,884 ms)
All Sites PING Time per PINGable Site	179.77 ms
All Sites Bytes per PINGable Site	43.61
All Sites Transmit Time per Byte	2.06 ms
Interstate Sites PING Time per PINGable Site	186.11 ms
Interstate Sites Bytes per PING	43.47
Interstate Sites Transmit Time per Byte	2.14 ms
Total Bytes Received per Session (All Sites)	513,221.40
Total Bytes Received from Interstate Sites	263,123.10
Estimated Percent of Pages Retrieved From Cache(s)	31.25%
Estimated Number LD Bytes Retrieved From Cache(s)	82,225.97
Number LD Bytes for Interstate Holding Time	180,897.10
Interstate Time for Receiving Web Pages	387,119.90 ms

The ratio of total session time (total holding time) to long distance holding time for pages received when browsing the WWW is: 387,119.0.1 ms / 3,994,884 ms = .0969 or 9.69 percent

This percentage does not exceed the 10% criteria of the mixed traffic rule and should be considered a worse case scenario based upon the characteristics of a majority of the selected subjects.

3.2.1 WWW Dialup Access Only

If we analyze only those sessions where dial-up access was used, we get approximately the same results as before from the 99 dial-up WWW sessions in the sample as shown below:

Variable	Average (Arithmetic Mean)
Number of Local Sites Visited	3.05
Number Long Distant (LD) Sites Visited	5.36
Number LD Intra-State Sites Visited	1.24
Number LD Inter-State Sites Visited	4.12
Number Pages Retrieved per Site	3.01
Number Bytes per Page from LD Sites	22,485.35
Number Bytes per Page from Local Sites	23,714.58
Time to View/Read All Pages	31.38 Min
Session Length (Total Holding Time)	72.19 Min (4,331,515.2 ms)
All Sites PING Time per PINGable Site	179.77 ms
All Sites Bytes per PINGable Site	43.61
All Sites Transmit Time per Byte	2.06 ms
Interstate Sites PING Time per PINGable Site	186.11 ms
Interstate Sites Bytes per PING	43.47
Interstate Sites Transmit Time per Byte	2.14 ms
Total Bytes Received per Session (All Sites)	598,323.48
Total Bytes Received from Interstate Sites	287,361.10
Estimated Percent of Pages Retrieved From Cache(s)	31.25%
Estimated Number LD Bytes Retrieved From Cache(s)	89,800.30
Number LD Bytes for Interstate Holding Time	197,560.69
Interstate Time for Receiving Web Pages	422,779.87 ms

The ratio of total session time (total holding time) to long distance holding time for pages received when browsing the WWW is: 422,779.87 ms / 4,331,515.2 ms = .0976 or 9.76 percent.

This percentage also does not exceed the 10% criteria of the mixed traffic rule.

3.2.2 WWW Direct Access Only

If we analyze only those sessions where direct access was used (DSL, TI, T3, etc.), we get approximately the same result as before from the 74 direct access WWW sessions in the sample as shown below:

Variable	Average (Arithmetic Mean)
Number of Local Sites Visited	2.07
Number Long Distant (LD) Sites Visited	4.44
Number LD Intra-State Sites Visited	.68
Number LD Inter-State Sites Visited	3.86
Number Pages Retrieved per Site	3.18
Number Bytes per Page from LD Sites	18,812.12
Number Bytes per Page from Local Sites	21,359.12
Time to View/Read All Pages	16.08 Min
Session Length (Total Holding Time)	58.97 Min (3,538,356 ms)
All Sites PING Time per PINGable Site	179.77 ms
All Sites Bytes per PINGable Site	43.61
All Sites Transmit Time per Byte	2.06 ms
Interstate Sites PING Time per PINGable Site	186.11 ms
Interstate Sites Bytes per PING	43.47
Interstate Sites Transmit Time per Byte	2.14 ms
Total Bytes Received per Session (All Sites)	405,765.00
Total Bytes Received from Interstate Sites	230,955.90
Estimated Percent of Pages Retrieved From Cache(s)	31.25%
Estimated Number LD Bytes Retrieved From Cache(s)	72,173.71
Number LD Bytes for Interstate Holding Time	150,782.20
Interstate Time for Receiving Web Pages	339,793.80 ms

The ratio of total session time (total holding time) to long distance holding time for pages received when browsing the WWW is: 339,793.8 ms / 3,538,356 ms = .0960 or 9.60 percent.

This percentage also does not exceed the 10% criteria of the mixed traffic rule.

If caching Web pages is ignored for Web searching, the ratio of Interstate holding time to total holding time (session time) is .141 = 563083.40 ms /3,994,884 ms or 14.1 % where 563083.4 = Total Interstate bytes transmitted * 2.14ms per byte or 263,123.1 * 2.14.

3.3 Summary of E-mail and WEB Browsing

The combined average of Interstate Long Distance holding time is (9.69 + 4.50)/2 = 7.095, but this is an unweighted average. Since e-mail is used 1.5 times as much as WWW searching and surfing according to several Internet surveys, a weighted average should be utilized. Thus, the weighted average is 6.57 = [(4.50* 1.5 + 9.69*1)/2.5], which is less than the 10% threshold. Thus, the data collected shows that a combination of e-mail and Web browsing account for less than 10% of Interstate traffic based on holding times as calculated.

Even if caching of Web pages is ignored for Web browsing, the combined average of Interstate Long Distance holding time is less than 10%, or (14.1 + 4.50)/2= 9.30 on an unweighted basis, or [(4.50*1.5) + (14.1*1)/2.5] = 8.34 on a weighted basis.

3.4 Internet Electronic Commerce

Electronic Commerce (EC) has two major categories of transactions, namely retail sales and trading partners. Trading partners utilize Electronic Data Interchange (EDI) which has characteristics similar to **e-mail** in that most of the time is consumed locally generating or processing an order or other business transaction with very little holding time required to transmit the data in the business transaction to its destination. It is estimated that EDI transactions will exhibit the same ratio of holding times as e-mail. A retail sale is more like a combination of web searching and e-mail. The shopping phase of a retail sale is like web searching (but with more reading time) but the buying phase is like e-mail in that it requires a form be completed (like composing a response to an e-mail message) and consumes very little time to transmit the relatively small amount of data involved. It is estimated that Electronic Commerce sessions will have holding time characteristics similar to the average of web browsing and e-mail and will utilize less than 10% for long distance Interstate holding time.

3.5 Internet Telephony and Fax

There are two forms of using the Internet for telephony. One requires a desktop at each end of the connection equipped with microphones and special hardware and software for digitizing voice. The entire end-to-end call is routed from the originating workstation to the local ISP, over the Internet to the destination ISP, and finally to the destination workstation. The second type of IP telephony utilizes the normal telephone handsets at each end for communicating but the call goes through special gateways at each end of the call with the middle portion being transported over the Internet. Internet telephony in this mode can be thought of as having three components:

1. A local call to the originator's ISP using the PSTN.
2. Routing to an IP-Telephony gateway, digitizing and compression of the voice signals, transport over the Internet for directory lookup,

connection to the destination IP-telephony gateway nearest the destination and transport of packets of digitized voice-using TCP/IP.
3. A local call from the ISP's IP-telephony gateway to the destination handset using the PSTN.

Although IP fax is a technical reality, there is growing evidence that e-mail attachments may provide a better technical solution that for many applications. In this case, the IP fax data should look similar to the e-mail data in this study. In any case, the sending of a fax over the Internet involves compression at the sender and decompression at the receiver. The majority of time is spent in the local loop with transmission time being a small portion of the total time similar to e-mail over the Internet.

3.6 Web Based Videoconferencing

Videoconferencing over the Web has a similar structure and uses mechanisms like Internet telephony. The system is comprised of two local calls and an Internet connection. The Internet connection appears to be 1/3 of the resources but based on compression techniques the actual holding time should be similar to IP Telephony. One difference is that a special computer may be designated as a reflector that is host to many videoconferences to which people can join. The user simply logs onto the reflector host. A reflector is not necessary for videoconferencing if the IP address of the other person is known but it is a one-to-one session. The major difference in this type of system is that digitizing, encoding and compression are most likely to occur at the user's site. In addition, systems may employ portal-based compression/packet handling schemes that would reduce the Internet transport requirements. This technology is in its infancy in terms of use and little empirical data is available.

3.7 Other Internet Applications

Database Processing

These applications fit Internet Web browsing characteristics in that a user provides search criteria or form-filling data to the Browser that sends the data to a server where a store or retrieval operation takes place. The server sends back a confirmation message to the user or a series of Web pages resulting from a search which the user must process. The majority of time is spent in the local look. As popular databases are mirrored through the world, there will be less need to use Interstate resources to access such databases.

Telnet

Telnet provides the capability to log onto an Internet server located anywhere in the world from a client machine. It makes the client appear to be a terminal connected to the host machine which communicate using the Telnet protocol. The Telnet protocol assumes that each end of the connection is a Virtual Terminal with a Virtual Keyboard and a Virtual Printer. As text is typed at the client, it is accumulated into a buffer and when a complete line of text has been typed (by pressing the Enter or Return Key), it is then transmitted to the destination. While the typing is being completed, no data is being transmitted and no transmission facilities are being used. If the text sent by the client is a command, then the host will execute the command and return the results to the client and await the next command to be transmitted. The user at the client end will spend much more time typing and reading the screen display than is spent transmitting data between the client and the host.

File Transfer Protocol (FTP)

The file transfer protocol (FTP) is used to transfer large amounts of data between a server and a client. This is often referred to as downloading or uploading files. FTP uses the client server model. Initially, a connection is made between a client and a server in command mode but when a command to upload or download a file is made, a second connection is opened called the data connection. The command link stays open for the entire session but the data link opens and closes at the start and end of each file transfer. The command link consumes very few transmission resources since the command are very short (GET, PUT, LS, etc.) and result in small amounts of data for transmission from the server to the client. Data link transfers are transmission intensive. In order to reduce time of transmission, most large files are compressed (Zipped) prior to transmission thereby reducing by 10 to 50 percent the file size and transmission time.

Internet Relay Chat (IRC)

This provides the capability for people all over the world to chat or talk with one another. It works using the client server model. The client software connects to a chat server where the user joins a channel. As anyone connected to the channel types a line of text, t is sent to the server which in turn sends it to others connected directly to the chat server or to other chat servers in the world closest to those logged into the channel. The servers are connected in a minimum spanning tree architecture so as to minimize transmission of messages on the Internet. Very popular chat servers are mirrored across many locations in the world. Most messages are short and consume very little transmission time compared to the composition and reading time. It is hypothesized that Chat sessions are similar to e-mail in terms of interstate and local holding times

Appendix A User Data Logs for E-Mail and WWW Browsing

Role of User:
Type of Organization:
ISP Vendor:
Speed of Connection in bits/second:
Date:
Country of Citizenship:
Type of Connection:

E-mail:
Time Session starts:
INBOX:
Number of 'local' messages:
Number of 'long distance' messages:
Number of instate: Number of out-of-state:
Average characters exclusive of attachments:
Number of messages with attachments:
Average attachments per message:
Types of attachments:
Average bytes of attachments:
RESPONSES TO E-MAIL:
Number of responses:
Average characters per response:
Number of 'local' responses:
Number of 'long distance' responses:
Average minutes composing a response:
ORIGINATING E-MAIL:
Number of messages composed:
Average characters in a message:
Average minutes composing a message:

Time session ends:
Time to get to a long distance call site:
PINGed long distance sites and recorded time:

Message	Site	Avg. Time (Ms)	Bytes/PING

Web Browsing:

Type of Connection:
Transmission Speed of Connection:
Purpose of the Session:
Time session starts:

Number of 'local' sites visited:
Number of 'long distance' sites visited:
Number of instate sites:
Number of out-of-state sites:
Average pages from a site:
Average bytes on a long distance site page:
Average bytes on a local site page:
Average minutes reading/viewing all pages at a site:

Time session ends:

PINGed long distance sites and recorded time:

Site	Site Identification	Average Time (Ms)	Bytes/PING